

(12)

AIR FORCE**PERFORMANCE MEASUREMENT REQUIREMENTS FOR
TACTICAL AIRCREW TRAINING**

Wayne L. Waag
Byron J. Pierce
Steve Fessler, Capt, USAF

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85240-6457

July 1987

Final Report for Period November 1983 - November 1986

Approved for public release; distribution is unlimited.

DTIC
ELECTE
AUG 28 1987
S D A

HUMAN RESOURCES**AD-A183 814****LABORATORY**

**AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601**

87 8 13 108

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

HAROLD G. JENSEN, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AD-A183814

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TR-86-62			5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Operations Training Division		6b. OFFICE SYMBOL (If applicable) AFHRL/OT		7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Williams Air Force Base, Arizona 85240-6457			7b. ADDRESS (City, State, and ZIP Code)	
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Air Force Human Resources Laboratory		8b. OFFICE SYMBOL (If applicable) HQ AFHRL		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601			10. SOURCE OF FUNDING NUMBERS	
			PROGRAM ELEMENT NO 62205F	PROJECT NO 1123
			TASK NO 35	WORK UNIT ACCESSION NO 02
11. TITLE (Include Security Classification) Performance Measurement Requirements for Tactical Aircrew Training				
12. PERSONAL AUTHOR(S) Waag, W.L.; Pierce, B.J.; Fessler, S.				
13a. TYPE OF REPORT Final		13b. TIME COVERED FROM NOV 83 TO NOV 86		14. DATE OF REPORT (Year, Month, Day) July 1987
				15. PAGE COUNT 42
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP		
05	08		aircrew performance measurement	
05	09		aircrew proficiency	
			aircrew proficiency evaluations	
			aircrew training	
			performance assessment	
			(Continued)	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)				
<p>This report attempts to identify Tactical Air Command's (TAC) needs for aircrew performance data and to evaluate current practices and procedures for obtaining such information. State-of-the-art measurement technology is reviewed and where considered appropriate, recommendations are made for near-term application. Finally, areas are identified requiring long-term research and development (R&D). Specific objectives addressed were:</p> <p>1. TAC's needs for performance measurement information necessary to support its training programs were defined. This required that consideration be given to both the user of measurement information and the intended use of such data.</p> <p>2. Criteria were defined whereby the adequacy of existing as well as potential measurement techniques could be evaluated. These criteria addressed the practical aspects of alternative measurement techniques as well as the more "scientific" aspects such as reliability and validity.</p> <p style="text-align: right;">(Continued)</p>				
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office			22b. TELEPHONE (Include Area Code) (512) 536-3877	22c. OFFICE SYMBOL AFHRL/TSR

Item 18 (Concluded):

performance measurement
performance monitoring
training evaluation
training management

Item 19 (Concluded):

3. Existing TAC measurement capabilities were reviewed and documented. This included a review of regulations and directives as well as the definition of specific techniques and practices. Also required was a definition of how specific techniques are being applied to meet specific measurement needs,

4. Existing measurement practices and capabilities were evaluated using the criteria previously defined. The intent was to determine how well current practices are meeting existing measurement needs and to identify specific areas which are in need of improvement,

5. State-of-the-art measurement technology and current R&D efforts were reviewed to determine where enhancements are possible and where existing solutions might be readily applied, *and*

6. Specific areas were identified that should be addressed in future R&D.



Accession For	
NTIS	<input checked="" type="checkbox"/>
DTIC	<input checked="" type="checkbox"/>
Other used	<input type="checkbox"/>
Classification	
Distribution/	
Availability Code	
Availability	Special
A-1	

SUMMARY

The purpose of this research effort was to define performance measurement requirements within the tactical training environment; to identify Tactical Air Command (TAC) needs for aircrew performance data; and to evaluate current practices and procedures for obtaining such information. Specific objectives addressed were:

1. TAC's needs for performance measurement information necessary to support its training programs.
2. Criteria whereby adequacy of existing as well as potential measurement techniques could be evaluated.
3. Existing TAC measurement capabilities, to include a review of regulations and directives as well as the definition of specific techniques and practices.
4. Measurement practices and capabilities, using criteria defined under the second objective to determine how well current practices meet existing measurement needs and identify areas needing improvement.
5. State-of-the-art measurement technology and current research and development (R&D) efforts, to determine where enhancements are possible.
6. Specific areas requiring future R&D.

To accomplish these objectives, information was gathered from three major sources: First, information concerning current procedures and capabilities, as well as user requirements, was obtained through structured interviews with TAC training and management personnel. Second, a review of performance measurement research literature and TAC regulations and directives was performed. Third, a review of current performance measurement technology was conducted in order to supplement the literature review. Analysis of this information enabled the six program objectives to be addressed. Overall conclusions were: (a) Performance measurement (PM) data currently available to support operational needs for performance monitoring, proficiency evaluation, and training management are adequate, although improvements in each area are needed. (b) PM data required for support of training evaluations are virtually nonexistent. (c) Current technology can be applied in certain areas and would result in definite improvements; however, such an approach may represent but a piecemeal solution to the broader problem of total training system design. (d) Further R&D is required--especially in development of objective performance measures for the area of training evaluation.

PREFACE

This research and development (R&D) was performed by the Operations Training Division of the Air Force Human Resources Laboratory. It supports Technical Planning Objective 3, Aircrew Training, the objective of which is to identify and demonstrate cost-effective strategies and new training systems to develop and maintain combat effectiveness. This effort was conducted under Work Unit 1123-35-01, Performance Measurement Requirements for Tactical Aircrew Training. The principal investigator was Dr. Wayne L. Waag.

The authors owe a debt of gratitude to personnel within Tactical Air Command (TAC) whose support made this research possible. Special thanks are due to Lt Col Mark Nataupsky, HQ TAC/DOTS, who served as the primary liaison between the Command and the Laboratory. It is through his efforts that the necessary coordination of the interviews at the individual units was made possible. Special thanks are also due to those individuals at the individual units who made the necessary arrangements and attended to the detailed scheduling of each interview. These included: Lt Col Mike Lackey, 405 Tactical Training Wing (TTW); Maj Wayne Warren, 58 TTW; Capt Kevin Court, 355 TTW; Maj Tom Horton, 474 Tactical Fighter Wing (TFW); and Maj Jim Peck, 49 TFW. And most importantly, thanks are extended to the individual aircrew and support personnel who provided the detailed information necessary to accomplish the research objectives.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
Problem Statement	1
Definitions	1
Background	2
Report Organization	3
II. PURPOSE OF THE EFFORT	3
Program Objectives	3
1. Definition of Measurement Needs	3
2. Definition of Evaluation Criteria	3
3. Documentation of Existing Capabilities	4
4. Identification of Deficiencies in Existing Capabilities	4
5. Application of Existing Technology	4
6. Identification of R&D Requirements	4
Program Scope	4
III. METHOD OF ACCOMPLISHMENT	5
User Interviews	5
Literature Review	6
Technology Review	6
IV. RESULTS AND DISCUSSION	6
Measurement Needs	6
Uses of Performance Measurement Data	6
Users of Performance Measurement Data	8
Operational Measurement Criteria	8
Definition Requirements	9
Validity Requirements	10
Reliability Requirements	10
Sensitivity Requirements	10
Practicality Requirements	11
Existing Measurement Capabilities	11
Regulations and Directives	11
Current Measurement Techniques and Support Capabilities	13
Application to Measurement Needs	15
Analysis of Current Capabilities and Procedures	16

Table of Contents (Continued)

	Page
Aircrew Performance Monitoring	16
Aircrew Proficiency Evaluation	18
Training Management	21
Training Evaluation	22
Application of Existing Technology	24
Aircrew Performance Monitoring	24
Aircrew Proficiency Evaluation	25
Training Management	26
Training Evaluation	27
Identification of R&D Requirements	27
Aircrew Performance Monitoring	27
Aircrew Proficiency Evaluation	28
Training Management	28
Training Evaluation	28
V. CONCLUSIONS	29
Overall Conclusions	29
Specifics	29
VI. RECOMMENDATIONS	31
REFERENCES	32
APPENDIX A: INTERVIEW PROTOCOL	35
APPENDIX B: INTERVIEW SCHEDULES	36

LIST OF TABLES

Table	Page
1 Uses and Users	9
2 Data to Meet Measurement Needs	15
3 Research Conclusions	30

PERFORMANCE MEASUREMENT REQUIREMENTS FOR TACTICAL AIRCREW TRAINING

I. INTRODUCTION

The purpose of this report is to define operational performance measurement requirements within the tactical training environment. An attempt will be made to identify Tactical Air Command (TAC) needs for aircrew performance data and to evaluate current practices and procedures for obtaining such information. State-of-the-art measurement technology will be reviewed and where considered appropriate, recommendations will be made for near-term application. Finally, areas will be identified requiring long-term research and development (R&D).

Problem Statement

The problem addressed in the present effort was identified in a Request for Personnel Research (RPR) entitled "Definition of Performance Measurement System Requirements for Tactical Aircrew Training" submitted by TAC to the Air Force Human Resources Laboratory (AFHRL). The RPR states:

Performance assessment techniques in tactical aircrew training programs have not incorporated recent technology improvements in performance measurement. In addition, present TAC aircrew performance measurement systems do not provide adequate discrimination between skill levels (both cognitive and psychomotor) of tactical aircrew members. Performance measurement on training devices and the mission equipment is not geared to a definition of mission readiness. Research has been requested and accomplished in a piecemeal fashion addressing only selected aircrew measurement problems. There has been no effort on the part of TAC or the research community to develop a comprehensive integrated research, development, and implementation plan for an aircrew performance measurement system. In summary, TAC has a two-part problem. First, the Command does not have an up-to-date comprehensive integrated aircrew performance measurement system. Second, the Command and the R&D community do not have an overall plan that defines the necessary actions to develop the required performance measurement system. This request is for the development of a plan for tactical aircrew performance measurement.

To summarize, the request was for a front-end analysis which would: (a) define the requirements for a comprehensive performance measurement system (PMS) that would provide TAC the necessary information for the conduct, management, test and evaluation of its training programs; (b) identify recent technological advances that could be readily applied; and (c) identify those areas requiring further R&D and develop a plan for its accomplishment.

Definitions

In order to ensure that the reader fully understands the terms in this report, some definitions will be presented. Perhaps the most basic term used throughout this report is "aircrew performance measurement." Simply stated, it is the act of ascertaining the dimensions, quantity, quality, etc. of the performance of aircrews by comparison against a standard. It is the application of a set of rules that categorize and/or quantify aircrew performance by comparison against a standard. To the extent that both the rules and standards are precisely

defined, the resulting measures are said to be objective. To the extent that such rules and standards are ill-defined, the resulting measures are said to be subjective. Examples of performance measures include training scores, bomb scores, number of ground-controlled approaches (GCA) per quarter, release parameters at the pickle point, written instructor pilot (IP) comments on gradeslips, opinions solicited during course evaluation visits, etc. In other words, performance measures are information or data used during the conduct, management, and evaluation of training. Thus, performance measurement within the context of this report refers to all information that is used in aircrew training; it is not synonymous with grades or similar judgments of an evaluative nature.

Two other terms are used extensively throughout this report: performance monitoring and performance assessment. Performance monitoring refers to the observation or recording of aircrew behavior. It is an essential ingredient of all training. The instructor must know what the aircrew is doing if he is to provide proper feedback. In some cases, the instructor can simply observe; in other cases, he must rely on a recording medium such as audio tape or film. Performance assessment goes one step further and introduces an evaluative component. An attempt is made to judge the "goodness" or "badness" of performance.

Although an attempt has been made to define these three terms independently, there is considerable overlap--especially between measurement and assessment. For this reason, the terms are treated much the same in the body of the report and are sometimes used interchangeably. For example, whether a grade is a measurement or an assessment is of importance from a conceptual viewpoint; practically, however, it makes no difference, and for this reason, the two terms are often used together.

Background

The importance of performance measurement has long been recognized within the Air Force flying training R&D community. Early work at the Air Force Personnel and Training Research Center (AFPTRC) in the 1950s focused on the development of objective scoring procedures for use in the T-6 aircraft. Performance record sheets were developed for in-flight use, in which the instructor was required to record specific events for each maneuver, such as maximum airspeed, altitude loss, etc. These performance record sheets were subsequently used to collect data on student performance, in an attempt to develop objective performance standards. Other efforts at AFPTRC focused on the use of motion pictures for recording of cockpit instruments during various flight maneuvers and on the use of such data to generate measures of performance.

The establishment of AFHRL in 1968 resulted in additional efforts aimed at the development of objective measures of pilot performance. By this time, computer technology had advanced to the point of allowing the rapid processing of large amounts of data. The capability to record objective flight parameters in both the flight simulator and the aircraft led to efforts to develop measures of pilot performance using fairly elaborate computational and statistical procedures. During the early and mid-1970s, research efforts at AFHRL focused on the development of measurement technologies for use as a tool in the Laboratory's ongoing R&D program (Waag, 1983). Most work centered on the development of measurement capabilities for the Flying Training Division's¹ research simulator, the Advanced Simulator for Undergraduate Pilot Training (ASUPT). At the time, the simulator was configured as a T-37B. A comprehensive measurement system was created to provide objective measures for representative maneuvers from all phases of T-37 training (Fuller, Waag, & Martin, 1980). Other work attempted to develop airborne measures of performance from data gathered from an instrumented T-37 aircraft (Waag & Knoop, 1977). When

¹Now the Operations Training Division.

the ASUPT was modified to an A-10 and F-16 configuration (and became known as the Advanced Simulator for Pilot Training (ASPT)), some of the measurement scenarios were modified, and new ones were developed. Again, the emphasis was on providing measurement tools for use in the ongoing simulator research program.

In the late 1970s, the need for improving measurement capabilities in the operational training environment was realized. An advanced development project was initiated to develop and evaluate integrated aircrew performance measurement systems for application to both flight simulators and aircraft. A prototype system was developed, implemented, and tested on a C-5 flight simulator (Waag & Hubbard, 1984). Although objective simulator and airborne performance measurement systems are an important R&D concern, they represent only a portion of the operational measurement problem. Aircrew measurement information can fulfill a number of operational training needs including: (a) determining present mission readiness levels; (b) predicting combat performance; (c) identifying personnel for training advancement; (d) providing feedback information to aircrew members concerning training progress; (e) validating training methods, procedures, and programs; (f) providing feedback to curriculum developers about the effectiveness/efficiency of training programs; and (g) providing information to training managers, who must decide how training resources are to be allocated. Presently, there is concern within TAC that information generated by current measurement techniques may be inadequate to meet some of these needs. As a result, the RPR that serves as the requirement for the present effort was initiated.

Report Organization

Section II of this report defines the purpose of the effort. It identifies, in detail, the program's objectives, its scope, and some of its limitations. Section III documents the methodology used and details the sources of, and the strategy for obtaining, the information used in the effort. Section IV presents the results and findings relevant to specific program objectives. Section V presents our conclusions; Section VI, our recommendations.

II. PURPOSE OF THE EFFORT

Program Objectives

The specific objectives of the present effort were as follows:

1. Definition of Measurement Needs

The first objective was to define TAC's needs for performance measurement information to support its training programs. This required that consideration be given to both the user of measurement information and the intended use of such data. The first task was to identify measurement needs within the following categories: (a) measurement needs for proficiency assessment and evaluation, (b) measurement needs for training development and evaluation, (c) measurement needs for training resource management, and (d) measurement needs for aircrew performance diagnosis. The second task was to identify all users of performance measurement information within the Command, ranging from the individual aircrew member within a squadron to headquarters personnel. In this manner, the information necessary to accomplish each intended "use" could be identified for each "user."

2. Definition of Evaluation Criteria

The second objective was to define those criteria whereby the adequacy of existing, as well as potential, measurement techniques could be evaluated. These criteria would address the

practical aspects of alternative measurement techniques, as well as the more "scientific" aspects such as reliability and validity.

3. Documentation of Existing Capabilities

The third objective was to review and document existing measurement capabilities within TAC. This included a review of regulations and directives, as well as the definition of specific techniques and practices and how they were being applied to meet specific information needs.

4. Identification of Deficiencies in Existing Capabilities

The fourth objective was to evaluate existing measurement practices and capabilities using the criteria defined under the second objective. The intent was to determine how well current practices were meeting existing measurement needs and to identify specific areas that were in need of improvement.

5. Application of Existing Technology

The fifth objective was to review state-of-the-art measurement technology and current R&D efforts, to determine where enhancements to address existing deficiencies were possible. The intent was to determine where off-the-shelf technology might be readily and effectively applied to these problems.

6. Identification of R&D Requirements

The final objective was to identify specific areas requiring R&D to address any identified deficiencies.

Program Scope

An attempt was made to make the scope of the present effort as broad as possible to ensure maximum generalizability. However, available resources limited the number of weapon systems selected for review. Factors considered were: (a) the number of fighter aircrews in initial and continuation training; (b) representativeness of the mission; (c) overlap of mission type, such as air superiority, ground attack, and multi-role; and (d) aircrew composition. The weapon systems jointly selected by AFHRL and HQ TAC/DOT were the F-15, F-16, and A-10. These were felt to be representative of TAC training programs. Moreover, the results should generalize across all TAC training programs because of program similarities dictated by regulation.

For these weapon systems, all missions and tasks necessary for effective operation were considered. The mission task listings were supplied by HQ TAC/DOT. No attempt was made to identify specific measurement techniques and shortfalls on a task-by-task basis, because of the extensiveness of these listings. Instead, only broad categories were addressed, such as air combat maneuvering, beyond-visual-range tactics, low-level navigation, emergency procedures, etc.

The present effort addresses all levels of training, with the exception of fighter lead-in training. Included are training conducted at the replacement training units (RTUs), continuation and upgrade training at operational units, and large-scale composite force exercises such as those conducted during Red Flag. Emphasis was placed on training at the RTUs and operational

units. All training media were addressed, including the aircraft, ranges, aircrew training devices (both simulators and part-task trainers), formal course academics, and ground training in support of initial and continuation training. Emphasis was placed on measurement capabilities for aircraft, ranges, and simulators. In addition to current simulation capabilities within TAC, other state-of-the-art systems were considered, such as the ASPT and the Simulator for Air-to-Air Combat (SAAC). Training support capabilities were also addressed. The Air Force Operational Readiness Management System (AFORMS) and the Cromemco microcomputers purchased by TAC were considered as well.

III. METHOD OF ACCOMPLISHMENT

This section describes the methods used to address the six program objectives. Information was gathered from three major sources. First, information concerning current procedures and capabilities and user requirements was obtained through structured interviews with TAC training and management personnel. Second, reviews were conducted of the performance measurement research literature, as well as regulations and directives within TAC. Third, a review of current performance measurement technology was conducted in order to supplement the literature review. This included both ongoing and planned R&D programs within the Department of Defense (DoD).

User Interviews

Structured interviews (see Appendix A) with TAC training and management personnel were the primary means of gathering information about the requirements for performance measurement data within training. The interview was designed to obtain the following information: (a) a description of the individual's job in terms of his/her duties and responsibilities; (b) a description of the data or information necessary to perform the job; (c) the availability of the information and any associated problems; (d) how the information was used, where it was sent, and in what form; (e) an indication of any perceived deficiencies or problems in the process; and (f) any recommended improvements.

As discussed in the previous section, one of the goals of the effort was to address all levels of training, as well as all individuals involved in the conduct and management of training. This required that individuals from all pertinent functions be interviewed at both the squadron and wing levels. Through coordination with HQ TAC/DOT, a list of such functions was compiled and used as the basis for identifying individuals to be interviewed. As a minimum, it was decided to gather data from at least one RTU and one operational unit for each of the weapon systems (A-10, F-15, and F-16). To conserve travel resources, the following wings were identified: the 405 Tactical Training Wing (TTW) and 49 Tactical Fighter Wing (TFW) for the F-15, the 58 TTW and 474 TFW for the F-16, and the 355 TTW and 354 TFW for the A-10. Because of the expected high degree of overlap among units in terms of their measurement requirements, it was decided to conduct detailed interviews at one RTU and one operational unit. The units selected for such detailed interviews were the 405 TTW and the 474 TFW. Interviews at the remaining units were oriented primarily toward identification of any differences. Interviews were also conducted with personnel at HQ TAC. Although it was originally intended that interviews at HQ TAC and the 354 TFW be conducted on the same trip, they could not be scheduled together; therefore, the visit to the 354 TFW was finally cancelled. The final list of offices interviewed at each location is presented in Appendix B. All visits were coordinated through HQ TAC/DOT.

Each interviewee was told the purpose of the effort and that it was being conducted at the request of HQ TAC. In general, the interviews followed the lines previously discussed. Upon completion of the interviews, summaries were prepared and forwarded to the respective squadrons to review for accuracy. Only minor changes were suggested.

Literature Review

The second major source of information included TAC regulations, directives, syllabi, etc., as well as the research literature. Documents from TAC that were reviewed included: (a) TAC Regulation (TACR) 50-31, Training Records and Performance Evaluations in Formal Flying Training Programs; (b) TACR 60-2, Aircrew Standardization/Evaluation Program; (c) Air Force Regulation (AFR) 50-38, Field Evaluation of Education and Training Programs, including the TAC supplement; (d) TAC course syllabi (B, TX, I) for the A-10, F-15, and F-16; (e) instructor/operator manuals for the A-10, F-15, and F-16 flight simulators; (f) the Air Combat Maneuvering Instrumentation (ACMI) user's manual; (g) local directives such as the 405th Wing Almanac, the 49 TFW Training Objectives Plan; and (h) informal documents such as Criterion Referenced Objective (CRO) listings, locally prepared simulator operators' handbooks, etc. Also reviewed were the various forms and reports used in TAC's training programs.

Literature reviewed from the research community included both AFHRL reports and those of other R&D organizations. Of special interest were reports that addressed some of the same issues addressed by the present effort. These included: (a) a requirements analysis for a combat-ready aircrew performance measurement system (Obermayer & Vreuls, 1974), (b) a requirements analysis for an F-16 performance measurement system (Schmid, Gibbons, Jacobs, Faust, & Moore, 1978), and (c) a requirements analysis for an air combat maneuvering performance measurement system (McGuinness, Forbes, & Rhoads, 1983). Reports attempting to develop measures for tactical tasks such as air combat maneuvering were also reviewed. A bibliography of the information reviewed is presented in Appendix E.

Technology Review

The third major source of information came from a review of current performance measurement technology with potential application to TAC training programs. It included R&D currently being conducted within DOD, as well as planned R&D. For all practical purposes, this consisted of research efforts within the Air Force and Navy. Included were efforts to develop improved performance measurement capabilities, efforts to develop training management support capabilities, and such generic technology improvements as microcomputer technology, computer graphics, data base management systems, etc.

IV. RESULTS AND DISCUSSION

This section contains the results of our investigation, presented according to the six program objectives described in Section II of this report.

Measurement Needs

The first objective of the present effort was to determine TAC's needs for performance measurement (PM) information. This entailed the identification of the uses of PM data as well as the users of such information.

Uses of Performance Measurement Data

From interviews conducted with squadron, wing, and headquarters personnel, current uses of PM information were identified within four broad categories: aircrew performance monitoring, aircrew proficiency evaluation, training management, and training evaluation.

Aircrew Performance Monitoring. If an aircrew is to make progress toward a desired goal, then performance monitoring of their current activities is required. In aircrew training, the instructor and/or evaluator must know how the aircrew is performing in order to give the necessary feedback and diagnosis. For example, if a pilot does not know where the bomb impacts, he cannot properly correct his own performance, and his error will likely continue. Operational performance feedback has been, is, and will continue to be a fundamental prerequisite for effective training. Performance monitoring is of concern in two primary domains: the simulator and the aircraft. In both cases, the requirement is the same: The instructor and/or evaluator must be given sufficient information about each performance to enable proper diagnosis of any performance deficiencies and suggest remedial action. The same requirement exists for day-to-day continuation training, where the individual must often judge and critique his own performance.

Aircrew Proficiency Evaluation. In the area of aircrew proficiency evaluation, there is a requirement for an evaluation of performance. Observed aircrew behavior is compared against a standard, which leads to an evaluative assessment of the acceptability of the behavior. Proficiency evaluations within TAC generally take the form of grades and are employed throughout all phases of training. Specific requirements include:

1. Training Progress. There is a requirement for the evaluation of aircrew performance on a mission-by-mission basis during formal training courses. This includes an assessment of overall mission performance, as well as performance on the individual mission elements. Such "grades" are required both for RTU courses and for operational unit courses leading to a particular qualification.

2. Recurring Qualifications. There is a requirement for recurring evaluations of aircrew performance as part of the Standardization-Evaluation Program prescribed by TACR 60-2.

3. Role Qualifications. Proficiency evaluations are required in order to determine the acceptability of aircrew performance for roles such as flight lead, mission lead, element lead, instructor, etc.

4. Mission-Ready Checks. There is a requirement to measure the "mission readiness" of units, as well as that of individual aircrews.

5. Local Area Checks. There is a requirement to evaluate the aircrew's knowledge of local area operations.

6. Continuation Training Event Requirements. There is a requirement for aircrews to perform those events outlined by TACM 51-50 for continuation training.

Training Management. Day-to-day training operations require the gathering and use of much information. Major PM information needs in the area of training management include:

1. Resource Scheduling and Requirements. Both personnel and equipment resources must be scheduled for duty. To do so requires the integration of a great deal of information from a variety of sources on a variety of tasks ranging from day-to-day scheduling of operations to the identification of long-term future resource requirements.

2. Recordkeeping. Recordkeeping requirements in TAC training are governed by regulations, headquarters requirements, and local directives. The types of data are quite varied and encompass the full range of activities within the training system. Generally speaking, if training information is generated, there is a requirement to record it.

3. Reporting. Similarly, if there is a requirement to record training information, there is usually a requirement to summarize it and report it. For virtually every function within the training system, there is a requirement to prepare some type of report to be forwarded to higher headquarters.

Training Evaluation. Finally, information is needed to evaluate and refine existing training programs and their various components. Because the function of training is to prepare aircrews for performance in the operational environment, there is a need for information that is indeed indicative of "mission readiness." Also, there is a requirement to assess the impacts that modifications to training programs have on aircrew proficiency. Training programs and their individual components, such as flight simulators, must be evaluated in terms of how well they prepare aircrews for mission performance. Any improvements at the RTU should be reflected by better individual aircrew performance at the gaining units. Likewise, other types of training such as Red Flag and the Aggressors should also lead to improved aircrew performance. Clearly, the evaluation of the effectiveness of such training programs requires performance measurement information.

Users of Performance Measurement Data

Virtually every individual within the training system has a need for some type of performance measurement data. It is likely that any user of such information will occasionally have a need for information from any of the four major use categories previously defined. However, most users will require information from only one or two of the categories for normal day-to-day operations. Based on the interviews with TAC personnel, the matrix shown in Table 1 was constructed to identify the major "users" of performance measurement information and the types of information that they require. Table 1 clearly shows that different users have differing needs for measurement information. The data most often associated with performance measurement, such as grades, are used primarily by those individuals directly involved in training (e.g., students, instructors, operational aircrews, and standardization-evaluation flight examiners (SEFEs)). Flight commanders, who represent first-level management, are also directly concerned with this type of data. Those units having a special training function (such as Red Flag, TAC ACES, etc.) have a need for aircrew performance monitoring data in order to improve performance feedback. Other personnel are primarily concerned with data relative to the supervision and management of training. It may be noted that only a few users require information relative to the evaluation of training. Again, it must be pointed out that there may be considerable consistency in terms of the needs of any one user, yet there will be instances when that individual requires data from the other categories. For example, even though the squadron operations officer is concerned primarily with data that reflect the day-to-day operations of the squadron, he must also be informed of substandard or poor performance demonstrated by aircrews in a syllabus training program.

Operational Measurement Criteria

The second objective of the effort was to define the criteria against which both existing and proposed performance measures techniques could be evaluated. As a result of the literature review, five criteria were identified: definition requirements, validity requirements, reliability requirements, sensitivity requirements, and suitability requirements. Each of these will be discussed in some detail.

Table 1. Uses and Users

	Performance monitoring	Proficiency evaluation	Training management	Training evaluation
RTU Units				
Student	X	X		
Instructor pilot	X	X		
Flight commander	X	X	X	
Standardization-evaluation	X	X		
WSTO			X	
Weapons	X		X	
Squadron CC/OPS			X	X
Training development				X
Scheduling			X	
Wing management			X	X
OPS Units				
Crews	X	X		
Instructor pilot	X	X		
Flight commander	X	X	X	
Squadron CC/OPS			X	X
WSTO			X	
Weapons	X		X	
Training			X	X
Scheduling			X	
Flight records			X	
Standardization-evaluation	X	X		
Intelligence			X	
Checked flag			X	
Wing management			X	X
Headquarters				
# A			X	X
HQ TA			X	X
Special Training Units/Exercises				
Red Flag	X	X		X
Aggressors	X	X		X
TAC ACES	X	X		X

Definition Requirements

Recall that one of the key elements of aircrew performance measurement is the comparison against a standard. The implication is that there must exist some criterion or standard metric against which the performance will be measured or compared; that is, there is a requirement to define precisely what constitutes a particular performance level. The requirement for a clear definition of the criterion is one of the basic assumptions that underlies all TAC training programs. As Schmid et al. (1978) pointed out, the criterion definition problem exists at two levels. That is, criteria must be defined at the individual task level (e.g., criteria for a particular level of performance of a barrel roll attack during air combat training) and they must also be defined at higher levels (e.g., criteria for designating the novice aircrew as "mission-ready.") Clearly, it is desirable that performance levels and their measures be defined as precisely and objectively as possible.

Validity Requirements

There is also a requirement that the measures themselves be valid; in other words, that they measure what they are supposed to measure. A number of types of validity are relevant. First, there is face validity. Do the measures appear to be valid? For the purpose of user acceptance, face validity is an essential requirement. Second, there is content validity. Do the measures address all of the essential components of task performance? For example, do bomb scores adequately assess all the essential components of weapons delivery? Third, there is predictive validity. Can the measures be used to successfully predict future events? For example, how well do the measures predict some future criterion such as mission readiness or combat effectiveness? Of these types of validity, the first two are probably more important in day-to-day training operations; however, the third is undoubtedly of extreme importance in the long term.

Reliability Requirements

The third requirement is that measures be reliable. In other words, the measures that result from any instance of aircrew performance should be consistent. To the extent that both performance standards and the rules for generating measures are well defined and, more importantly, that these rules are followed, the resulting reliability will be quite high. Two types of reliability are of concern within training programs. The first is inter-rater reliability. If two instructors observe the same performance, will their assessment be the same? To what extent are grading practices standardized? The second is intra-rater reliability. How consistent is the individual instructor in his/her assignment of grades? Will equivalent performances always produce the same assessments? Reliability is an extremely important criterion, because without reliability, there can be no validity.

Sensitivity Requirements

A fourth requirement is that measures be sufficiently sensitive for their intended purpose. For example, you would not use a bathroom scale to weigh a letter to determine the amount of postage required. The scale would not be sufficiently sensitive for the intended purpose. There are analogous situations in aircrew training.

Basically, there are four types or levels of measurement. The most rudimentary is the nominal level, which merely classifies events into categories. For example, counting the occurrences of different types of errors during a flight represents a nominal level of measurement. Counting events necessary to meet TACM 51-50 requirements is also an example of nominal-level measurement. There is no underlying dimension or continuum. The next measurement class is the ordinal level, in which categories are rank-ordered along some underlying continuum or dimension. Flight grades are a good example of measurement at the ordinal level. In this instance, the underlying dimension is aircrew proficiency. A "1" is better than a "0," a "2" is better than a "1," etc., but the difference in proficiency between a "0" and a "1" is greater than between a "2" and a "3." At the interval level, it is assumed that there is an equal distance between the categories. For example, if flight grades were at an interval level of measurement, then the difference in proficiency between a "0" and a "1" would be the same as the difference between a "2" and a "3." At the highest level are ratio scales, in which there is a true zero point. Circular error during weapons delivery is a good example of a ratio scale.

The importance of levels of measurement lies in the intended use of the information. There can be instances when data are not sufficiently sensitive for their intended use. For example, an evaluation of the effectiveness of a training device requires that performance data be

gathered and statistically analyzed. Clearly such data as TACM 51-50 event counts would not be sufficiently sensitive to detect any differences in performance since all aircrews generally perform the same number. Likewise, the use of grades may also not be sufficiently sensitive in discriminating among performances, since the vast majority of marks are either a "1" or a "2." Because virtually everyone receives the same grades, it would be impossible to detect real differences in performance.

Practicality Requirements

Finally, performance measures must be practical. Perhaps the most important consideration, at least from the user's standpoint, is simplicity. Measures must be easy to obtain, easy to interpret, easy to record, etc. This is especially true of data gathered during airborne missions. Although techniques are available for gathering very detailed information using the instructor as an observer, their impact on instructor workload and safety-of-flight considerations makes them operationally unsuitable. In some instances, the data may not be available. For example, knowing what the aircrew is looking at throughout the mission could provide very powerful diagnostic information; however, such data are not currently available. Moreover, the costs associated with the instrumentation to gather such information may not be justified in terms of its actual benefits. The cost issue is definitely a factor to be considered in the implementation of objective types of scoring systems.

Existing Measurement Capabilities

This section of the report will describe the existing measurement capabilities within TAC and discuss their application to identified user needs.

Regulations and Directives

To a large extent, the uses of performance measurement data in day-to-day training operations are governed by regulations and directives. The major regulations include: TACR 50-31, which is concerned with grading; TACR 60-2, which is concerned with the Standardization-Evaluation Program; TACM 51-50, which is concerned with training event requirements; AFR 50-38, which concerns the field evaluation of training programs; and local supplements and directives, which serve to implement the above regulations. Each of these will be discussed below.

TACR 50-31, Training Records and Performance Evaluations in Formal Flying Training Programs. This regulation governs the basic grading and reporting process during formal training. It defines the grading criteria for both the overall mission and mission elements, as well as course training standards. It also identifies the forms to be used for the documentation of training, as well as procedures for their completion. Current grading criteria are as follows:

Grade	Explanation of Grade
Unknown	Performance not observed or element not performed
Dangerous	Performance was unsafe (one element marked "dangerous" requires overall grade of "0")
0	Performance indicates lack of ability or knowledge
1	Performance is safe, but indicates limited proficiency. Makes errors of omission or commission.
2	Performance is essentially correct. Recognizes and corrects errors.
3	Performance is correct, efficient, skillful and without hesitation.
4	Performance reflects an unusually high degree of ability.

It should be noted that the regulation also allows for specification of additional criteria such as written behavioral objectives; however, such criteria are not mandatory.

TACR 60-2, Aircrew Standardization-Evaluation Program. This nine-volume regulation governs TAC's Standardization-Evaluation Program. Volume I describes the organization of the program and how it is implemented, the grading policies and systems used during the conduct of evaluations, the procedures for reporting the results of evaluations, and the trend analysis program for isolating common problems. It also establishes policies for written examinations. For overall qualification levels, there are three grades that can be assigned: Exceptionally Qualified (EQ), Qualified (Q), and Unqualified (U). For each area examined during the evaluation, there are also three grades that can be assigned: Q, which represents the desired level of qualification; Q-, which indicates that the examinee is qualified but needs additional training or corrective action; and U, which indicates unqualified.

In the remaining volumes of the regulation, an attempt is made to link these grades to criteria. For instrument evaluations, an attempt is made to specify tolerances for specific flight parameters. For example, altitude deviations for a steep turn must be within a tolerance of + or - 200 feet to receive a grade of Q. For tactical evaluations, the criteria are usually not as precisely defined. For example, a grade of Q for offensive maneuvering requires "effective use of basic fighter maneuvering and air combat maneuvering to attack/counter opposing aircraft. Good aircraft control. Effectively managed energy levels during engagements." A grade of Q- reflects "limited maneuvering proficiency; however, during engagements did not effectively counter opposing aircraft. Occasionally mismanaged energy levels resulting in the loss of an offensive advantage." It is clear that a great deal of subjective judgment may enter into these assessments, because the criteria are loosely defined.

TACM 51-50, Flying Training. This 10-volume manual identifies training requirements at all levels, including initial qualification training, mission qualification training, continuation training, and specialized training. Of interest with regard to performance measurement is Volume VI, which gives the required number of events per half (6 months) for each area. For example, to maintain a mission-ready status, each F-15 aircrew must achieve at least one deployable automatic relay terminal (DART) hit each 6 months.

AFR 50-38, Field Evaluation of Education and Training Programs. This Air Force regulation and the TAC supplement govern the evaluation of TAC's formal training courses. There are two major sources of evaluation data: graduate evaluation forms and field trips. After an initial 90-day period with the squadron, the new aircrew is evaluated by the assigned instructor, using the graduate evaluation form. The form consists of a listing of tasks performed by the new aircrew during mission qualification training. The new aircrew is rated according to how well each task is performed. Completed forms are returned to the RTU, where the results are summarized and trend analyses performed. A report is produced semiannually. Periodically, a training development team visits the operational units. Interviews are conducted with various squadron personnel for the purpose of identifying common problems that RTU graduates are having. Upon completion of these field trips, the results are summarized, and a report is produced and forwarded to higher headquarters. These two vehicles are the primary means of providing an external evaluation of the training program.

Local Directives. At the unit level, local supplements and directives tailor the previously discussed regulations to the unique needs of the squadron and wing. For example, there is always a local supplement to Chapter 7 of TACR 60-2, Volume I, concerned with local procedures. There are other documents that govern wing operation. For example, the Operations Order for the 405 TFW or Luke Almanac (F-15) gives the detailed operational procedures for both scheduling and training. Similar documents are also found at the operational units.

Current Measurement Techniques and Support Capabilities

This section describes TAC's current measurement techniques and support capabilities. These descriptions are based on the previously described regulations and directives, as well as information gathered from the interviews.

Event Accomplishment. Event accomplishment is perhaps the most widely used measurement technique in TAC training. Either a "yes" (performed successfully) or a "no" (not performed successfully) provides information that is used for virtually all training functions. Such data are used primarily for training management. In continuation training, maintaining mission-ready (MR) status is based on event accomplishment.

Written Examinations. Written examinations are employed in all phases of training. There are published standards that indicate the required score to achieve a "passing" mark. These examinations consist of objectively scored items which can be marked either "correct" or "incorrect." Item analyses are periodically performed to ensure proper test reliability. A variant of the written examination is the oral examination, such as the recitation of flight manual (Dash-1) procedures. In all instances, the intent is to test the aircrew's knowledge of a specific domain of information.

Grades. As was previously discussed, grading is governed by regulation--specifically, TACR 50-31 and TACR 60-2--using the same scales throughout TAC training programs. For standardization-evaluation checks, the criteria are established in the regulations themselves. For routine grading, some discretion is allowed in terms of additional criteria, such as behavioral objectives or criterion-referenced objectives (CROs). This has resulted in some differences among RTUs. For example, the 405 TTW (F-15) publishes CROs for the various courses it teaches; the 58 TTW (F-16) does not.

Objective Performance Data. Some objective performance data that are available in TAC training programs include:

1. Instructor Observations. Perhaps the greatest source of objective data is the observation of aircrew behavior by the instructor. A requirement of the grading system is that the instructor provide a written commentary regarding any grade of "1" or less. Generally, these comments provide an objective description of the errors committed by the aircrew. Such commentary is the primary means of conveying information to a subsequent instructor as to particular strengths or weaknesses of a particular student. In such instances, the instructor is acting as a "recorder" of performance, rather than an "evaluator" of performance. In the past, researchers have successfully used such techniques as a means of gathering objective airborne performance data.

2. Simulator Measurement Data. Objective performance data are also available on student performance in the simulators. Because such devices are usually digitally driven, an abundance of performance data can be extracted. Objective data are available at two levels. First, "raw" data can be displayed at the instructor station. Such digital information is akin to that obtained from repeater instruments and is used for performance monitoring. Second, objective "measurement" capabilities are also available. These include weapons scoring and a parameters-monitoring capability for the A-10, F-15, and F-16 Operational Flight Trainers (OFTs); a procedures-monitoring capability for the A-10 and F-16; and an approach-monitoring capability for the

F-15. Although such objective "measurement" capabilities are available, they tend not to be used, with the exception of weapons scoring.

3. Weapons Delivery Scores. Both bomb scoring and strafe scoring capabilities are available at the conventional ranges. The strafe score indicates the number of hits per pass. The bomb score indicates circular error, as well as clock position of the impact point. For tactical ranges, a system is available at some locations which attempts to measure impact point through triangulation techniques using television cameras.

4. Airborne Audio and Video Recordings. Both audio and video techniques are used to record airborne performance data. Audio recorders are used by instructors as a means of aiding mission reconstruction for the purpose of debriefing. Radar, heads-up display (HUD), and gun camera film are also available for recording key events and phases during airborne missions. Most recently, video recorders have become available, so that the tapes can be replayed immediately after completion of the mission. These recording techniques are all aimed at providing mission feedback for debriefing. They are also used as a means of assessing shot accuracy for air-to-air engagements. For ground attack missions, they are useful for indicating the actual parameters at weapons release, measures which cannot be obtained from range scores.

5. Air Combat Maneuvering Instrumentation (ACMI) Data. The ACMI provides a means of recording airborne engagements for the purpose of creating a ground replay for later debriefing. The ACMI consists of four major subsystems: the airborne instrumentation subsystem, the tracking instrumentation subsystem, the control and computation system, and the display and debriefing system. The ACMI is capable of measuring and recording a relatively large number of objective flight parameters. The data are then displayed in alphanumeric and graphic format for real-time observation and post-mission replay and debriefing. Real-time missile flyout models are available to provide weapons scoring information.

6. Electric Warfare Range Instrumentation Data. The Tactical Fighter Weapons Center (TFWC) Range Complex (TRC) at Nellis AFB provides a means of generating and recording objective performance data for electronic combat (EC) events (e.g., threat engagements). The majority of terminal threat simulators and emitter systems have been instrumented to provide time-tagged digital information concerning threat system status and aircraft position. Such data are superimposed on the threat optics and recorded on tape. These tapes can then be used by threat analysts for evaluation and for debriefing of aircrews. In addition, emitter-receiver processors (ERPs) have a capability for providing radar imagery and generating a missile miss-distance or a gun score. There is also a prototype Surface-to-Air Threat Assessment (SATA) system, which provides a missile/gun flyout based on engagement data collected from the Range Instrumentation System (RIS). The RIS gathers and combines aircraft and threat data from a number of sources to enable a replay of an engagement using SATA.

7. Time. One objective measure used throughout TAC training programs is time. For each mission, time flown is logged and used for a variety of purposes. Time could also be used to document other performances, such as time on the boom during aerial refueling, time in a particular weapons envelope, etc.

Opinion Data. Another type of information used extensively throughout TAC's training programs is opinion data. A prime example is interview data, similar to the kind of data gathered during the conduct of the present effort. Field trips which are part of the graduate evaluation program gather such information. Open-ended questionnaires such as course critiques are another example. Likewise, summary reports such as the TAC Form 134 require a written summary evaluation of the aircrew's training performance. A major characteristic of such information is that it is qualitative in nature. It generally requires that the provider of the

information make some type of summary judgment based on experience. This does not imply that such information is invalid or incorrect; however, it does place certain limitations on the use of such information for some of the measurement needs that have been identified.

Accident/Incident Data. Another type of information used is accident/incident data. The occurrence of an accident/incident results in the generation of a report. These reports are summarized and trend analyses are performed at both HQ TAC and the Safety Center at Norton AFB, California. Such information could be considered a specific type of event data.

Data Automation. Data automation does not represent a measurement technique but rather, a support capability. Computer support is provided at three levels. First, support is provided by the local base computer facilities. Primarily, this includes routine recordkeeping functions for maintenance and supply. Second, there is an Air Force Operational Readiness Management System (AFORMS). This system, which currently makes use of base computing facilities, is concerned primarily with recordkeeping for TACM 51-50 event requirements. Third, TAC has procured microcomputer systems for use in the individual squadrons; however, actual usage depends on the individual squadron's requirements.

Application to Measurement Needs

This section describes those measurement techniques currently being used to support each of the four types of measurement needs. Based on the information gathered from the squadron interviews, Table 2 was constructed to depict these relationships. In the following paragraphs, the measurement data supporting each need will be discussed.

Table 2. Data to Meet Measurement Needs

	Performance monitoring	Proficiency evaluation	Training management	Training evaluation
Event accomplishment	X	X	X	
Written examinations		X		
Grades		X		X
Objective data				
Instructor pilot observation	X			
Simulator data	X			
Weapons scores	X	X		
Recordings	X			
Air Combat Maneuvering Instrumentation Data	X			
Electronic warfare range data	X			
Time	X	X	X	
Opinion data				
Interviews				X
Questionnaires				X
Reports		X	X	X
Accident data				X
Data automation		X	X	

Aircrew Performance Monitoring. From Table 2, it is clear that only objective data are used in the area of performance monitoring. The observation of aircrew performance for the purpose of providing effective feedback necessitates the availability of objective information.

Aircrew Proficiency Evaluation. A fairly large number of measures are used to assess aircrew proficiency. Objectively scored measures include event accomplishment, written examinations, weapons scores, and time; however, the most widely used measures are grades. Opinion data in the form of reports are also used for summary evaluations, such as TAC Form 134. There is some data automation support for the proficiency evaluation function. In some squadrons, for example, the Cromemco microcomputers are used for the calculation of Top Gun scores, as well as for providing summary statistics for weapons delivery scores. Similarly, AFORMS is used to record and summarize continuation training event accomplishment.

Training Management. The day-to-day management of training is largely based on objective data in the form of event accomplishment and time. Such requirements as syllabi, TACM 51-50 events, unit test equipment (UTE) rates, Programmed Flying Training (PFT) student flow rates, standardization-evaluation checks, etc. drive day-to-day training operations. To a large degree, the primary job of both squadron- and wing-level management is to ensure that all of these requirements are being achieved. Scheduling is the function whereby these requirements are translated into day-to-day operations. At present, there is some data automation support for training management; however, it is extremely limited, especially for the scheduling function. AFORMS is used to record and summarize TACM 51-50 events requirements, and as such, its products are used to some degree in the training management function. The Cromemco microcomputers have also been employed to some limited degree for long-range resource forecasting.

Training Evaluation. As previously discussed, the evaluation of TAC's formal courses is governed by AFR 50-38. The primary means for performing an external validation of training are through the graduate evaluation forms that are sent to the gaining units, and through the field trips. Thus, the primary data used for evaluation are grades and opinion. Internal validation of the training program is accomplished primarily through course critiques and an analysis of failed missions. The analysis of failed missions is based on the grades for the individual mission elements.

Analysis of Current Capabilities and Procedures

This section of the report attempts to apply the operational measurement criteria to existing techniques and capabilities in order to determine how well the four measurement needs are being satisfied. Discussion of each measurement need will begin with an overall assessment (i.e., opinion of the authors), followed by an identification of specific deficiencies.

Aircrew Performance Monitoring

Information necessary to monitor aircrew performance is of importance primarily for the feedback it provides. It is necessary to know "what" the aircrew is doing both in the air and in the simulator. As discussed in the previous section, the data that provide such information are objective in nature. The two primary criteria used for evaluating this information were: (a) its completeness (i.e., content validity), and (b) its suitability (i.e., availability, cost). Overall, current performance-monitoring capabilities work reasonably well. Generally speaking, there is at least some information available to enable a reasonable reconstruction of airborne missions, at least for the critical mission elements. Audio and video recordings, as well as direct observation, are the primary means used for missions other than those flown on the ACMI or

electronic warfare (EW) ranges. For simulator training, there is generally sufficient information at the instructor operator station (IOS) console to enable a reasonable monitoring of performance. Although the current techniques "work," limitations exist, primarily in the area of information completeness. The following sections will discuss areas where there are definite information gaps which limit the amount of performance feedback.

On-Board Monitoring Capabilities. As mentioned, the primary on-board capabilities are audio and video recordings. Although audio recorders are used widely, they are not embedded within the aircraft. This clearly represents an equipment limitation. Also, HUD camera and gun camera film are in use in some squadrons, but because of the lengthy delay (normally 1 day) in the processing of the films, their usefulness for debriefing is extremely limited. Videotape recorders (VTRs), which are generally becoming available, overcome the limitation of the lengthy processing time; however, they were not currently available in all of the units surveyed. This was particularly evident at the RTUs, where the requirement for information feedback is perhaps the greatest. Even VTRs have certain limitations, however. First, their recording time is limited to 30 minutes. This necessitates turning the recorder on and off throughout the mission. Second, there is no provision for a split-screen capability, which would permit simultaneous replay of both the radar display and HUD. Third, there is a rapid-access problem in terms of difficulty in locating the part of the mission that the instructor wishes to review. Fourth, for some tasks (such as EC), there is no recording of data other than the audio from the Radar Warning Receiver (RWR).

Aside from these limitations associated with current VTR technology, there is the more important problem of information completeness. Although the data recorded on the VTR are extremely useful for debriefing, they do not present the "big picture." This is especially true for air combat maneuvering (ACM) engagements. VTR replays are useful primarily for the terminal phases of the engagement, including shot assessment. However, the maneuvering phases of the engagement are very difficult to reconstruct. This usually requires a reconstruction from memory, with the aid of comments from the audio recording. Although reconstruction of a 1 versus 1 engagement may be fairly accurate, attempting to reconstruct larger engagements becomes increasingly difficult.

Range Capabilities. One possible solution to the problem of information completeness are the instrumented ranges. Clearly, an advantage of flying missions on the ACMI is the graphic replay, which enables an accurate reconstruction of the engagement. However, the ACMI is not without its own limitations. First, it too provides insufficient information. An area where the insufficiency of information is perhaps greatest is the lack of a "trigger squeeze" downlink capability on the F-15 and F-16. This has led to manual shot insertion by the Range Training Officer (RTO), which is prone to much error. In a study by Hooks and Kress (1984), it was found that approximately 33% of all shot assessments were in error, due to insertion delays by the RTO. Moreover, the study did not address delays or possible omissions by the aircrew in terms of calling their shots. Clearly, this severely impacts the accuracy of the weapons scoring on the ACMI. In fact, for missions flown on the ACMI, it was observed that shot assessments are still based on the HUD/gun camera film or VTR recordings. Another area of concern is the lack of radar lock-on data. It should be noted that fixes are being developed for these problems.

A second and perhaps more important limitation of instrumented ranges is range access. To begin with, the number of ACMI ranges is fairly limited; thus, not all units have access to the ACMI. Even for those squadrons located near an ACMI, there can still be problems of access. For example, one unit investigated in the present effort has access but only on a very limited basis. The priority for the range's use as a tool for routine unit training is very near the bottom. Similarly, another unit was found to have difficulty gaining access, due to higher

priority given to another Government agency. At yet another unit, all DART training is accomplished on the ACMI range, thus limiting the amount of access for other purposes.

Some of the limitations of the ACMI are even more severe for the EW ranges. Perhaps the greatest limitation is access. The number of ranges is extremely small; hence, access is extremely limited. In addition, the utility of data on aircrew EC performance is limited. Many of the existing threat simulators are insensitive to the use of countermeasures. Even in the case of ERPs, which are sensitive to countermeasures, the scores generated are limited by the lack of highly accurate aircraft position information. As mentioned, the SATA system is a prototype which provides a missile/gun flyout capability and computes the closest-point-of-approach of the weapon to the aircraft. It depends on aircraft track and threat data from other sources. As a debriefing tool, it is not currently useful because of the lengthy delays in processing. For example, the processing of flight path and firing event data requires several hours. In addition, the current threat models are sensitive only to gross electronic countermeasures effects (e.g., radar break-lock). Thus, the system is very limited as a means of providing aircrew performance monitoring and feedback.

Simulator Capabilities. Because the simulators for the weapon systems reviewed are digital, there is no shortage of information that is computed and made available for display. In all cases, there is a remote IOS console which displays data of interest. Each console contains a number of cathode-ray tubes (CRTs), repeaters, annunciators, controls, etc. In general, the limitations are not in terms of the availability of the performance data but rather, the means by which the data are displayed. Much information is presented digitally and as such, is often difficult for the instructor to interpret--especially in those situations in which there is an analog representation in the actual aircraft. Other problems include: too much information being presented in a single display, information being presented in too many different locations, and information being presented in a manner that is inconsistent with its presentation in the cockpit.

These problems were observed in all the simulators investigated. For example, in the F-16 OFT, cockpit activity can be monitored only by selecting a physical section of the cockpit. Unfortunately, procedures do not necessarily relate to the specific areas of the cockpit. Therefore, it is quite likely that the IP may miss some important activity. The simulator does have a procedures monitoring capability. However, it is not up to date; so, the feature is not used. In the A-10 OFT, there is a procedures monitoring capability, but it is also out of date; and hence, not used. In the F-15 OFT, the monitoring of procedures requires information from mixed sources, including repeater instruments, annunciator lights, and graphic representations of the cockpit. Final approach information is limited to a gauge which displays degrees left/right of centerline and feet above/below the glidepath. From the gauge, it is very difficult to determine trend information. These are but representative examples of the types of problems with IOS displays. Again, the problem is not the availability of information within the simulation but rather, how it is organized and displayed at the console.

Aircrew Proficiency Evaluation

The second major need lies in the area of aircrew proficiency evaluation (i.e., how the observed aircrew performance compares with the standard). There are two requirements: first, that the standards or criteria be well defined; second, that the rules for translating observed performance into an "assessment" or "grade" be well defined. For all practical purposes, a yes-no judgment is required. Does performance meet the standard? Grading systems in both training and standardization-evaluation virtually have a dichotomous outcome, despite the number of points on the scale. The current systems do "work," in that they are able to differentiate between acceptable versus unacceptable aircrew performance; however, despite the fact that they

work, there are limitations--both in the definition of performance standards and the grading process itself. In the following sections, these limitations will be discussed.

Definition of Performance Standards. Recall that a major requirement for any measurement is that the standard be well defined. Certainly, such a requirement is well accepted for the measurement of physical dimensions and quantities. However, for the measurement of behavior and in particular, aircrew performance, its importance is often not emphasized. In principle, TAC recognizes the need for defining standards, as evidenced by their inclusion in TACRs 50-31 and 60-2. In practice, however, their actual definitions are often very imprecise and subject to interpretation by the examiner or evaluator. Attempts have been made by some units to specify CROs to assist in grading; however, not all units have adopted such CROs. Moreover, the existence of CROs is no guarantee that they are being used or that they are being used in a standardized manner. For example, some IPs reported that they did not use CROs at all; others reported that they used CROs frequently. Still others reported that CROs were used only when a grading decision was not "obvious"; that is, when it was difficult to judge whether a performance should be graded a "1" or a "2." In no instances were standards appropriate to other points of the grading scale.

Overall, our observations indicated that TAC's grading and evaluation standards were not precisely defined in most instances, although exceptions exist such as qualification standards for weapons delivery, parameter deviations for instrument flight, etc. The general impression was that standards of "acceptable" performance are primarily a judgment call by the instructor or examiner. Without objectively defined standards of performance, proficiency becomes a matter of individual judgment, and such judgments can vary from individual to individual (i.e., different instructors have differing internal standards) and from time to time (e.g., an instructor's standards of acceptable performance may change as a function of experience as an instructor).

Despite the fact that standards were seldom found to be well defined, there do exist practices and procedures which tend to minimize the impact of this condition. There are numerous "checks" within the training system to safeguard against the bias of any one individual. For example, RTU students fly with more than one instructor. Such a practice not only exposes the student to different instructors who use different techniques, but also provides a safeguard against the possible bias of one individual. There also occur numerous "cross-talk" sessions in which standards are discussed. Moreover, one of the primary goals of the standardization-evaluation program is to promote standardized instruction and assessment. Nonetheless, there still exists the need for better-defined standards at all levels of training. Despite the "checks" just mentioned, which tend to safeguard the training system as a whole, failure to define standards precisely has its greatest impact on one key element: grading.

Grading. Grades are required by regulation for all levels of training. Generally speaking, each mission element, as well as the overall mission, must be graded. The syllabus dictates which events require the demonstration of proficiency (i.e., a "2" or better) for each mission. Because grades require the comparison of observed aircrew performance against a standard, how meaningful are they when this comparison is made against poorly defined performance standards? Four criteria for evaluating grades (performance measures) are reliability, validity, sensitivity, and practicality. Recall that in practice, grading is used to make a yes-no decision; that is, did the aircrew's performance meet the required standard? Grades are sufficiently sensitive for such a yes-no decision. They certainly meet practicality requirements in that they are easy to obtain, affordable, etc. The reliability of grades is probably acceptable as well, because of the yes-no nature of the measurement and the fact that the syllabus dictates required proficiency levels on a mission element basis. For example, there are instances in which a particular mission element must be graded a "2" on the third sortie within a phase in order to be satisfactory. Most likely, the aircrew will be given a "1" on the first sortie, either a "1" or a "2" on the second, and a "2" on the third. In such instances, any computed measure of reliability would be quite high.

The fourth criterion is validity. If an instructor assigns a "2," is the performance really a "2"? If the performance standards are ill defined, the validity of such a grade must be questioned. The opinion that the meaning of grades is questionable was also expressed by those instructors and students interviewed. In fact, the consensus was that grades in terms of the numbers were of little value and represented little more than a "square-filling" exercise. The following practices point to the variability that can be associated with any given grade:

- As mentioned previously, CROs, when they exist, are used (or not used) quite differently among different instructors. In other words, instructors do not use the same standards.

- Variation occurs in terms of "when" the grade sheets are completed. Some are completed soon after the mission, during the debrief; in other instances, they are completed much later, when the IP has the available time. As the time lapse increases, the accuracy of recall is likely to decrease, thus affecting the assigned marks.

- Grades can be used merely as a means of "documenting" poor performance or they can be used as the basis for the elimination of aircrews from training.

- Grades are sometimes more reflective of syllabus requirements than actual aircrew performance. For example, proficiency rides require an overall grade of "2" as well as all missions elements being a "2." In the event the aircrew performs poorly on a single element (i.e., a "1"), the instructor is reluctant to grade it as a "1" since this would require that the mission be reflown.

All of these examples point to the questionable validity of the current grading system in terms of whether the marks assigned are reflective of actual aircrew performance. In most instances, the yes-no decisions are probably correct; however, in many cases they are not, for one reason or another. Nonetheless, the current system does "work" for the purpose of differentiating acceptable and unacceptable aircrew performance at a fairly gross level. As pointed out during the interviews, perhaps the greatest value of grading is that it requires the instructor to consider each mission element. If the grade sheet were merely a "blank piece of paper," important aspects of mission performance might be overlooked during the debrief.

The current TAC grading system is described in a report by Martin (1984), which contrasted the athletic training model with current flying training practices. The report had this to say about ratings:

The 0 to 3 rating scale, as everyone who uses it knows, is really a 2-point scale, since the extremes are rarely used. It has the appearance of an objective scale, since the definitions of each category refer to criterion-referenced objectives (CROs) for each task. However, the CROs have never been worked out for many tasks and are not used for many others. The end result of this type of measurement is that it is useless. It cannot be used to discriminate between differences in skill levels except at the low-end extreme. It cannot be used to track progress. It does not function as a motivational tool since everybody gets the same mark. It cannot be used to assess the effects of changes in the flying program. This lack of sensitivity may have indirectly resulted in the loss of many flying hours in the mid to late 70s.

Measurement and Assessment. A corollary limitation is the lack of objective measurement and assessment, with the exception of weapons scoring information. This is partially due to the lack of objectively written performance standards. It is also due to the relatively high cost of

generating objective data in both the simulation and airborne environments. Nonetheless, it is possible to implement objective scoring capabilities for a wide range of mission tasks. The advantage of objectivity in scoring is that it "forces" the precise definition of performance measures in terms that a computer will "understand"; that is, stated in quantifiable physical units. Certainly, it must be admitted that there are many areas of aircrew training for which validated objective measurement techniques do not exist. In those instances, R&D is necessary to develop and validate objective measures of aircrew performance. The lack of objective measurement capabilities will be discussed further in the section dealing with training evaluation, where the impact is considered greatest.

Training Management

As mentioned earlier, the day-to-day management of training is based largely on objective data in the form of event accomplishment and time. The major elements of the training management function include tracking, recordkeeping, reporting, and resource scheduling. Despite the fact that data requirements for this function are fairly extensive, there is generally little difficulty in terms of either availability or quality of information, largely due to the fact that most of the required data take the form of event accomplishment. Thus, measurement techniques per se do not pose a problem. However, the manipulation and application of such information are very time-consuming and labor-intensive since there is very little computational support for training management functions. At present, the management of aircrew training within TAC can best be described as a manual paper-and-pencil activity. Although some support does exist from AFORMS and the Cromemco microcomputers for these functions, it must be characterized as very limited in scope. The following sections will discuss what are considered to be the primary problems in the area of training management; namely, the lack of computer support for many functions, and the limitations of such support where it does exist.

Data Automation Support. As mentioned, training management in TAC is primarily a pencil-and-paper exercise. The amount of available data automation dedicated to this use is very limited. Thus, manpower resources required for these management functions are quite significant. Usually, these functions are performed as additional aircrew duties. As such, they are not viewed very favorably within the units, since they detract from the primary mission of flying, and usually result in very long duty days. Perhaps the most demanding of these functions is in the area of scheduling. At present, from two to four individuals are assigned to be schedulers within a squadron. The scheduling task remains a manual paper-and-pencil and greaseboard activity. There exists no computer assistance for short-term scheduling operations, although the Cromemco has been used to a limited degree to support long-range forecasting. However, it is the weekly and daily scheduling operations that are the most labor-intensive. Clearly, computer assistance in this area represents a pressing need and could achieve significant manpower savings.

Tracking, recordkeeping and reporting are also primarily paper-and-pencil activities. There exists no computer-based management information system for routinely performing these functions, with the exception of those performed by AFORMS as will be discussed in the next section. Thus, recordkeeping and reporting are very labor-intensive activities. We observed many instances of multiple tracking of the same information. Often this is done to provide a data source for correcting errors. For example, an important management objective is to ensure that the proper number of hours are flown during each 6-month period. For each mission, the aircrew reports on AF Form 369 the number of hours flown. The sum of hours flown from all sorties must equal the allocated number of hours, as determined by the UTE rate. Inevitably, discrepancies occur which require a manual search to determine where the error lies. Personnel must ensure that the number of hours actually flown corresponds to the number of hours that should be flown. Clearly, the

requirement for such precision and the resulting manpower needed to ensure that the numbers "agree" must be examined. A similar situation exists with the tracking of TACM 51-50 events and, for that matter, most things that are to be recorded.

The lack of a computer-based management information system has its greatest effect at the RTUs, where the data requirements are the largest. Again, all tracking, ranging from sortie accomplishment to the analysis of failed missions, is done manually. The lack of computer support also affects the training development teams. For example, all item analyses of academic tests are accomplished manually, as are all quarterly reports and trend analyses. Clearly, there is a need for computational support for these functions. What is required is a comprehensive management information system that addresses the host of management functions found within a wing and its component squadrons. In addition to the tracking, recordkeeping, and reporting functions, it should also provide capabilities to support both near-term and long-range scheduling.

Implementation of AFORMS. At the time the present effort was being accomplished, AFORMS was in the process of being introduced to all units within TAC. Some of the units were already using AFORMS, whereas others were awaiting its implementation at the time of the interviews. The purpose of AFORMS is to serve as a replacement for Tactical Air Force Training Management System (TAFTRAMS) and provide a vehicle for tracking a variety of data, the most important of which relate to TACM 51-50 events. For the operational units in particular, one of the most important requirements is an up-to-date record of these event accomplishments. This is particularly true toward the end of each 6-month period, when there is usually a flurry of activity to ensure that all training requirements have been met for each aircrew (if not, the aircrew will lose mission-ready status). The major complaint expressed was that neither AFORMS nor the older TAFTRAMS system could provide an up-to-date record of event accomplishment. Generally, there is a 1- to 2-day delay in data entry to the AFORMS system. Some interviewees perceived this to be a management problem rather than a limitation of the system. It is possible that all flights completed by 1300 hours on one day could be input into the system by the following morning. Although lengthy delays are usually not critical early in each 6-month timeframe, they become important toward the end of the period. For the future, there are plans to have direct entry of information into the data base via optical character readers so that the data base can be updated in near-real time.

There are also problems in the readability and interpretability of the reports that are currently generated. At present, these reports tend not to be very "user-friendly." Moreover, it is usually difficult to perform "non-standard" retrievals of information from the data base. However, planned refinements should alleviate these problems.

In sum, AFORMS might be described as an emerging and somewhat immature system. For those problems that were identified, fixes have been planned. However, in addition to the technical problems associated with the system, those interviewed cited other shortcomings in this area, such as the need for increased training in AFORMS use and greater management emphasis at the unit level.

Training Evaluation

The fourth major need for performance measurement data lies in the area of training evaluation. The purpose of training is to prepare aircrews to perform during combat; in other words, to win the war. In peacetime, however, the next best criterion is "combat readiness." If the current training system and its individual components are to be properly evaluated, they must be evaluated against the criterion of combat readiness; and there must be measures available to do so. Examples of training system components that require such measures for determining their value

in terms of combat readiness include: full-mission simulators including TAC's OFTs, as well as special devices such as the Simulator for Air-to-Air Combat (SAAC); range instrumentation systems such as the ACMI; specialized training exercises such as Red Flag or the Aggressors; formal training courses; and perhaps most importantly, training in the aircraft itself. The underlying assumption in each of these cases is that the training provided should in some measurable way improve the combat capabilities of the Tactical Air Forces (TAF). Because each of these training capabilities is quite expensive, there is the inevitable question: Does the improvement in combat readiness offset the cost? Clearly, performance measurement data are needed to answer these questions.

Unfortunately, there is little information currently available which can be used to address such issues. The only types of evaluative data routinely gathered are grades and opinions, which may or may not reflect the later performance of graduates from the RTUs. These assessments are based on the students' performance during Mission Qualification training and as such, are not intended to be measures of combat readiness. There currently exists no means of quantifying operational performance. There are no measures currently available which can be used to quantify "how good" an individual aircrew performs. The lack of data for use in training evaluation is considered to be a major shortcoming of the current training system. In the next few sections, some of the specific limitations will be discussed.

Criteria for Combat Readiness. At present, there exist no accepted criteria or performance standards for combat readiness. It is assumed that if an operational aircrew meets their TACM 51-50 event requirements and passes their recurring standardization-evaluation checks, then by definition they are "combat-ready." Despite the fact that the appropriate squares are filled, it is clear that there are wide differences in readiness among aircrews. The lack of clearly defined criteria and performance standards was viewed to be a major problem--the solution of which is not a trivial matter. Until this problem is addressed, the situation will remain much as it is today, where evaluations of training and its components are based on little more than subjective opinions.

Despite the need and potential uses of such information for evaluative purposes, there still exist strong negative attitudes toward objective measurement and assessment, for it is feared that such data might be misused or used inappropriately. Current assessment techniques (standardization-evaluation, grades, events) are so structured that everyone receives the same marks; therefore, they cannot be used to discriminate among aircrews for the purpose of appraising their performance. The availability of quantifiable measures would provide that capability; however, implementation of such a capability would require strong Command commitment and possibly dramatic changes in policy.

Objective Measures. A closely related problem is the lack of objective measures for evaluating training on individual tasks. At present, objective measures are available for only a few tasks, such as weapons delivery. The lack of objective measures was identified as a problem in the area of proficiency evaluation. It is also a problem in the area of training evaluation. Although grades at least "work" for proficiency evaluation purposes, they lack sufficient sensitivity to be useful for training evaluations. For example, attempts have been made to use the TAC grading system for the conduct of transfer-of-training studies. The inevitable result has been a lack of any distinguishable differences, regardless of the training treatment. In the test and evaluation of the A-10 OFT, results showed that grades did not provide sufficient sensitivity to be useful measures (Rasinski, Rabeni, Holck, & Pierce, 1983). It was recommended that an R&D program be initiated to develop an objective measurement capability. Also, a study conducted some years ago at AFHRL demonstrated positive transfer from weapons delivery training in a T-37 to performance in an F-5 (Gray & Fuller, 1977). Although objective measures in terms of qualification criteria and circular error showed significant transfer, IP ratings produced "no

differences." Clearly, the lack of objective measures, as demonstrated in these efforts, is a major problem in the area of training evaluation. Readily available information in the form of grades or standardization-evaluation checks is simply inadequate.

Training Information Data Base. In an ideal sense, training evaluation represents a continual process. It is a crucial part of Instructional System Development (ISD), since it provides information whereby the training system is refined and updated. At present, there exists no training information data base that could be used as a vehicle for continually validating and refining TAC's training programs. The closest approximation is the tracking system for TACM 51-50 event requirements, AFORMS; however, since each aircrew must accomplish the same events, it does not provide the discriminative information necessary for use in training evaluation and validation. Currently, the only objective data that could be used for the establishment of a training information data base are weapons delivery scores and missile data. Although such data do not represent a complete range of performance information, they would provide the vehicle for initially establishing such an information data base. As more objective information became available, the data base could be expanded.

Application of Existing Technology

The previous section identified deficiencies and limitations in current performance measurement capabilities. This section discusses possible improvements using existing technology, in terms of the four major categories of operational measurement needs.

Aircrew Performance Monitoring

Limitations in performance monitoring capabilities were shown to exist in three environments: on-board the aircraft, on the ranges, and in simulators. Generally speaking, on-board limitations relate to information that can be monitored and/or recorded. Most of the problems associated with airborne video recordings can be addressed with existing technology. Longer recording times are available. Also, a split-screen capability is within the state of the art, as are mechanisms for setting "flags" during a mission to aid in rapid access of the desired portion of the sortie. Rapid advances in video technology are expected to result in systems having even greater resolution, longer recording times, smaller physical dimensions, etc.

Two problems associated with the ACMI were the lack of trigger squeeze information and the lack of radar lock-on data. Again, solutions to these problems are well within the state of the art. "Fixes" to the problem have been identified and are currently being tested. An additional limitation of the ACMI is the lack of accurate altitude information at low levels. These low-level data are necessary for the reconstruction of air-to-surface missions. The range at Nellis AFB is currently being expanded to provide a low-level capability. The Red Flag Measurement and Debriefing System (RFMDS), which is similar to the Navy's EC range located at Fallon, will have the capability to monitor, record, and replay missions flown in an interactive EC environment. Although the changes just discussed are technologically feasible, they do not address the basic limitation of all ranges: access. There still remains a unit-level requirement for systems that will enable the recording and replay of missions.

Problems in the display of information at the IOS can also be addressed with existing technology. The technology exists to display virtually any type of information or data computed within the simulation, in any type of format, whether it be digital or graphic. Current-generation graphics systems are quite powerful and can easily support any requirements for a dynamic, high-resolution display. The problems that remain lie not in the ability of the

hardware to support the display of information but rather, in the design and formatting of the data to be presented.

Aircrew Proficiency Evaluation

Limitations in proficiency evaluation capabilities exist in three areas: performance standards, grading, and objective measurement and assessment techniques. The current grading system shows a need for clearly defined performance standards and standardized grading practices; however, it does serve to differentiate between acceptable and unacceptable aircrew performance. Moreover, we found no other grading systems that would lead to any dramatic improvements. For example, the Navy also uses a four-point grading system, but it is perhaps even less stringently defined than TAC's grading system.

Some consideration was also given to the use of expanded rating scales, such as those used in certain R&D applications. However, the requirement to accurately define the standards associated with each point on the scale was viewed as operationally unsuitable. As mentioned, current standards and CROs, where they exist, are applicable to the "2" level of the current grading system. Given the difficulty in clearly defining the standards for a single level, expanding the scale would merely compound the problem.

Thus, no major changes to the grading system are recommended. However, greater emphasis should be given to two areas. First, an attempt should be made to more clearly define performance standards on a task-by-task basis. Where CROs do not exist, they should be developed and used. Second, an attempt should be made to improve the standardization within the current grading system. The role of CROs within the training system must be clearly defined and their usage in day-to-day training operations standardized. These recommended changes do not represent any new technology; rather, they are applications of well-established training principles.

The remaining problem area concerns the lack of objective measurement and assessment techniques for evaluating aircrew proficiency in simulators, on ranges, and aboard the aircraft. Over the past few years, a significant amount of R&D has been accomplished in developing objective measurement capabilities for flight simulators. Much of the technology is fairly mature and considered suitable for transition to operational use. One of the first devices to contain objective measurement capabilities was the Automated Flight Training System, which provided instruction for certain instrument procedures including GCAs and air intercepts. This system was eventually added to all F-4 and A-7 simulators within the Air Force. Also, a stand-alone instructional support system (ISS) has been developed and implemented on the F-14 OFT at NAS Miramar. The ISS contains a measurement capability whereby performance can be evaluated against predetermined criteria. This capability includes both continuous parameter monitoring and procedures monitoring. Based on curriculum and instructor inputs, grading criteria are provided for each task module (i.e., training objective). Once a task module is completed, the instructor is provided immediate feedback concerning the aircrew's score and any errors committed. A measurement system has also been developed for the C-5 flight simulator that is capable of providing objective indices of performance for such tasks as instrument procedures, checklists, normal and emergency procedures, and navigational profiles. Although the system was developed for a transport application, the tasks measured have much in common with those currently trained in TAC's OFTs. Most recently, an objective measurement capability was developed for the Strategic Air Command's (SAC's) B-52 aerial refueling part-task trainer located at Castle AFB. The capabilities of these systems are described in a report by Waag (1987).

Based on the capabilities of these prototype systems, an instructional support system design handbook was recently completed for use by Major Commands, who must identify functional

requirements, and the Simulator System Program Office, who writes specifications (Easter et al., 1986). The handbook will address not only performance measurement capabilities, but the entire range of instructional support technology (features such as record/playback, freeze, mission generation, etc.). Again, the handbook's concern is with the entire instructional support system of the simulator, of which performance measurement is but a single component. The important point is that the design of performance measurement capabilities for simulators should not be considered in isolation, but must be considered within the overall context of the instructional support capabilities of the system, since many of the individual features are interrelated. Included within the handbook is a sample specification for a somewhat generic F-16 OFT.

Training Management

In the area of training management, the two major limitations are the immaturity of AFORMS and the limited amount of data automation support in the areas of tracking, recordkeeping, reporting, and resource scheduling. For the most part, these problem areas can be addressed using existing technology. Certainly, computer-based management information systems have been used for many years in a variety of applications. For example, many of the reporting and recordkeeping functions within the RTUs are similar in scope to those within Air Training Command (ATC). Their Base Management System (BMS), which has been in existence for quite a few years, keeps track of all missions flown, as well as the individual mission element grades. From this data base, a variety of reports are prepared which are used for tracking the progress of each class and its individual students. The system employs an optical scanner which "reads" each grade sheet. A tape is prepared each evening, consisting of records for all missions flown that day. By the next morning, the data base is updated, and the reports incorporating the previous day's information are prepared for distribution.

ATC has conducted a test and evaluation of its Time-Related Instructional Management (TRIM) system at Laughlin AFB. The system is designed to include computer-assisted instruction (CAI) capabilities, as well as resource management capabilities. The resource management system (RMS) part of TRIM was designed to be a replacement for the BMS. However, it also provides an automated flight scheduling function. As such, it tracks the progress of each student for all training events such as academics, simulators, and aircraft sorties. Based on the availability and priorities of students, IPs, and resources, an automated scheduling routine generates a candidate schedule. The scheduler can then manually override as desired, to produce an updated schedule. Upon completion of the sortie, the IP completes the grade form, which is then entered directly into the data base via an optical scanner. All recordkeeping and reporting functions are accomplished by the RMS, thereby eliminating much paper-and-pencil activity.

Clearly, the most difficult aspect of the entire training management function is resource scheduling. Aside from the current ATC system, other demonstrations have occurred. In February 1982, AFHRL demonstrated a daily flight scheduling capability at the 479 TTW at Holloman AFB. The Forward-Looking Resource Scheduling (FLRS) system represented an application of technology developed at AFHRL as part of the Advanced Instructional System. Utilizing student, IP, course syllabus, and schedule data bases, the FLRS system assisted the scheduler by producing a "first-cut" basic schedule that was syllabus-specific and conflict-free. The scheduler was then able to fine-tune the schedule through the use of an on-line CRT. Unfortunately, FLRS required a large mainframe computer system for operation. Nonetheless, it did demonstrate the feasibility of computer-assisted scheduling operations. Efforts are currently underway to explore the applicability of FLRS to E-3 training.

The Navy has developed a computer-based management information system called the Aviation Training Support System (ATSS). It supports all the recordkeeping and reporting functions

previously described and also has a provision for automated flight scheduling. For example, at NAS Miramar, it tracks all students in approximately 50 different courses, including aircrews within the Replacement Air Group (RAG). ATSS also has a recordkeeping function for maintenance support.

From these examples, it should be clear that existing technology is sufficient to support TAC's needs in the area of training management--especially for tracking, recordkeeping, and reporting functions. Although computer-assisted flight scheduling has been successfully demonstrated, its applicability to TAC operations, especially at the RTUs, is unknown due to the complexity of such operations.

Training Evaluation

Three major limitations were identified in the area of training evaluation: the lack of criteria for combat readiness, the limited number of objective measures for use in training evaluation, and the lack of a training information data base that could be used for the continual refinement and update of the training system. Clearly, the development of a training information data base is well within the state of the art, as discussed in the previous section on training management. However, a review of the literature and existing technology revealed little that could be readily implemented in the other two areas. For example, Mixon (1982) and Mixon and Moroney (1982) have compiled comprehensive annotated bibliographies of research studies that have either developed or used objective measures of aircrew performance. Of the several hundred articles reviewed, none was concerned with criteria for combat readiness. Moreover, few addressed the development of objective measures for tactical combat tasks such as air combat maneuvering, air intercept, ground attack, etc. Clearly, these areas require further R&D.

Identification of R&D Requirements

The previous section identified existing technology that could readily be implemented. This final section identifies areas where solutions are not yet available and further R&D is necessary. Again, these will be discussed in terms of the four major categories of operational measurement needs.

Aircrew Performance Monitoring

Two R&D requirements were identified in the area of aircrew performance monitoring. First, there is a need to pursue the development of a unit-level airborne monitoring and debriefing capability. Although airborne recordings using either film cameras or VTRs provide much information that can be used for debriefing, they fail to provide a reconstruction of the "big picture." Ranges can provide this capability; however, there are simply not enough of them to support the required unit-level training. Even where ranges do exist, there is no guarantee that they are readily available. R&D is needed to explore alternative technologies for providing an airborne monitoring and debriefing capability at the unit level. One alternative that has been proposed is the use of flight data recorders as a means of gathering airborne information. Such data from different aircraft could then be multiplexed, and a graphic replay similar to the ACMI could be created. The primary benefit of such a system would be for unit-level training. For large-scale exercises, ranges such as ACMI would still be required.

Second, there is a need for the development of improved graphic techniques for information display. Such techniques would be of use for on-line monitoring and debriefing of missions flown

in both the aircraft and the simulator. As discussed, the capabilities of graphic systems have expanded greatly. The limiting factor is usually the design of the information display. R&D is needed to explore various graphic presentations for use across the range of tactical missions. For example, what is the best way to graphically depict aircrew performance in a high-activity electronic combat environment? What information is required by the instructor in order to properly monitor such combat situations? What types of display options should be available to the instructor teaching ACM?

Aircrew Proficiency Evaluation

Additional R&D is needed to develop and validate alternative approaches to the evaluation of aircrew proficiency. There are serious shortcomings that severely limit the usefulness of the data (grades) that are generated. Alternative scoring approaches should be explored which rely more heavily on the direct observations of aircrew behavior by the instructor. What is needed is a scoring system based directly on instructor observations rather than subjective interpretation or assessment of those observations in the form of a "1" or "2." R&D is also needed to develop techniques for quantifying those multidimensional higher-order concepts that are used throughout training, such as aircrew discipline, situation awareness, airmanship, and aggressiveness. Although such concepts are routinely used in training, they are not presently amenable to quantification. There currently exist no standardized definitions or means of measuring these concepts, despite the fact that they are considered quite important in the operational training community.

Training Management

Very little R&D appears necessary in this area. As mentioned, computer-based management information systems that are well within the state of the art will adequately handle the tasks of tracking, recordkeeping, and reporting. A possible exception is in the area of resource or flight scheduling. Although a number of systems have already been demonstrated, it is unknown whether there would be a direct application to the scheduling problem within TAC, especially at the RTUs. It is expected that further work on FLRS would address this issue.

Training Evaluation

Two R&D requirements were identified in the area of training evaluation. First, there is a need for the development and validation of objective task performance measures for use in conducting training evaluations. Emphasis should be placed on tactical war-fighting skills including both air-to-air and ground attack components. There is a need for the development of air-to-air measures for both close-in ACM and beyond visual range (BVR) tactics. For ground attack, emphasis should be placed on measures of aircrew performance in a high-threat electronic combat environment. The measurement system must also address the aircrew's effective use of sensor-based systems such as Low-Altitude Navigation and Targeting Infrared Night System (LANTIRN). Ultimately, what is required is a full-mission performance measurement system that provides assessments across the spectrum of tactical combat tasks. In addition to the development and validation of such a measurement capability, consideration must also be given to its eventual implementation. The most likely candidates for eventual implementation include the ranges (ACMI and RFMDS) and airborne data recording systems. For example, in the area of aircrew performance monitoring, a need was identified for a unit-level recording and debriefing facility. Obviously, consideration should be given for the inclusion of measurement capabilities within the ground-based processing system in addition to the replay capability. Clearly, these development efforts need to be coordinated and integrated.

The second R&D requirement is for the development of criteria and measures of combat readiness. Since combat readiness to some degree involves an aircrew's proficiency on individual tasks, such an effort must await the development of task-specific measures. However, combat readiness encompasses more than the performance of individual aircrews. Crews are trained to fight as units. Clearly, criteria and measures must be developed to describe the performance of units such as two-ship or four-ship operations. However, even the performance of such units may not be indicative of combat readiness in any meaningful sense. The true test is whether a squadron or wing is able to successfully execute its primary mission plan. To do so requires not only that the individual aircrews or fighting units perform adequately, but also that there be adequate logistics and maintenance support to enable mission accomplishment. Clearly, the concept of combat readiness is multidimensional in nature and must include all those elements necessary for mission success. The development of criteria and measures of combat readiness is viewed as a long-range R&D goal.

V. CONCLUSIONS

The present effort identified four operational needs for performance measurement information: (a) for monitoring the performance of aircrews in the simulator and in the air for the purpose of providing training feedback, (b) for assessing the proficiency of aircrews to ensure that training objectives are being achieved, (c) for managing training operations on a day-to-day basis, and (d) for evaluating the effectiveness of the training system and its various components. Major questions to be answered included: (a) how well current capabilities meet these needs, (b) where existing technology might be readily applied, and (c) where additional R&D is necessary. Although each of these questions has been dealt with in some detail, some "bottom-line" conclusions are presented below. These are summarized in Table 3.

Overall Conclusions

1. Performance measurement (PM) data currently available to support operational needs for performance monitoring, proficiency evaluation, and training management are adequate in the sense that they "work"; however, there are serious limitations and deficiencies in each area.
2. In the area of training evaluation, the PM data required for support are virtually nonexistent. Readily available data in the form of grades are inadequate to meet this need.
3. Current technology can be applied in certain specific areas and would result in definite improvements. However, generally, such an approach may represent but a piecemeal solution to the broader problem of total training system design. The specific information required in the areas of proficiency evaluation, training management, and training evaluation is very much dependent upon the structure of the training system. The application of "new" technology to support an "old" training concept may be a waste of resources.
4. Further R&D is required--especially in the development of objective performance measures for use in the area of training evaluation.

Specifics

1. There are limitations in the performance monitoring capabilities for simulators, ranges, and aircraft. The most serious deficit is the lack of a recording and debriefing facility for units that do not have easy access to either ACMI or EW ranges.

Table 3. Research Conclusions

Area	Deficiency	Solution
Aircrew Performance Monitoring	- limited on-board audio/video monitoring capability	existing technology
	- limited information displayed on IOS console	existing technology and R&D
	- limited data available from ranges	existing technology and R&D
	- lack of unit-level debriefing capability	R&D
Aircrew Proficiency Evaluation	- lack of well-defined training performance standards	existing technology
	- inadequate grading procedures	R&D
	- lack of objective measures and assessment capabilities	existing technology & R&D
Training Management	- limited data automation support	existing technology
	- lack of resource scheduling capability	R&D
	- immaturity of AFORMS	existing technology
Training Evaluation	- lack of criteria to measure combat readiness	R&D
	- limited number of objective measures	R&D
	- lack of a training information data base	existing technology

2. Well-defined criteria and performance standards and standardization in their use represent critical needs for all phases of training. This includes measures for individual tasks as well as for more global measurement areas such as situation awareness, aggressiveness, combat readiness, etc.

3. Only a limited number of objective measures are available to support aircrew proficiency evaluation. There are virtually no objective data for use in support of training evaluation.

4. Grades are of little value for conducting a training evaluation, due to their lack of sensitivity. Their validity in day-to-day training operations is also questionable, due to the lack of well-defined criteria and non-standardized usage.

5. At present, the management of training is primarily a paper-and-pencil activity. There is insufficient data automation support for tracking, recordkeeping, reporting, and resource scheduling. The problem is particularly critical at the RTUs. The use of aircrews to perform many of these manual functions is considered a waste of valuable resources.

6. AFORMS must be considered an emerging system. As such, it still has a number of problems which remain to be solved as the system matures. Given sufficient training in its use and increased management support, AFORMS can become a very useful tool for training management within the operational units.

VI. RECOMMENDATIONS

Based on our findings, we believe the following actions are necessary to develop a measurement capability to meet TAC's current and future requirements:

1. Apply existing principles and technology as follows:

a. Retrofit HUD/gun cameras with VTRs, including a split-screen capability where feasible. The minimum recording time should be 60 minutes.

b. Complete the implementation of a trigger squeeze and radar lock-on downlink capability on the ACMI for the F-15 and F-16.

c. Define training objectives or CROs clearly and standardize their usage within the training system.

d. Implement objective performance measurement capabilities in TAC's OFTs where possible.

e. Implement a computer-based management information system within RTUs. The system should include the capability of flight scheduling. It should also be expandable to include performance information from gaining units.

2. Support R&D in the following areas:

a. Development of task-specific measurement techniques for air-to-air (ACM/BVR) and tactical ground attack in a high-threat EC environment.

b. Development of techniques for measuring concepts such as situation awareness, aggressiveness, etc.

c. Development of improved graphic techniques for information display in all training environments (aircraft, ranges, and simulators).

d. Development of unit-level airborne monitoring and debriefing capabilities for both air-to-air and ground attack missions.

e. Development of criteria and techniques for assessing combat readiness at both the individual and unit levels.

f. Development of scoring approaches based on direct observations of aircrew behavior.

REFERENCES

- AF Regulation 50-38 (1981, July). Field evaluation of education and training programs. Washington, DC: Department of the Air Force.
- Easter, A.W., Kryway, J.T., Olson, W.R., Peters, S.M., Slemon, G.K., & Obermayer, R.W. (1986, May). Development of instructor support future guidelines (AFHRL-TR-85-57(II), AD-A168 285). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- Fuller, J.H., Waag, W.L., & Martin, E.L. (1980, August). Advanced simulator for pilot training: Design of automated performance measurement system (AFHRL-TR-79-57, AD-A088 855). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- Gray, T.H., & Fuller, R.R. (1977, July). Effects of simulator training and platform motion on air-to-surface weapons delivery training (AFHRL-TR-77-29, AD-A043 649). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.
- Hooks, J.T., & Kress, G. (1984, September). Application of air combat maneuvering instrumentation (ACMI) data for performance measurement (AFHRL-TP-83-36, AD-B086 064L). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- Martin, E.L. (1984, October). Practice makes perfect (AFHRL-TP-84-32, AD-A147 124). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- McGuinness, J., Forbes, J.M., & Rhoads, J.E. (1984, March). Air combat maneuvering performance measurement system design (AFHRL-TP-83-56, AD-B081 060L). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- Mixon, T.R. (1982). An annotated bibliography of objective pilot performance measures: Part II (NPS55-81-010PR). Monterey, CA: Naval Postgraduate School.
- Mixon, T.R., & Moroney, W.F. (1982). An annotated bibliography of objective pilot performance measures (NAVTRAEQUIPCEN IH-330). Orlando, FL: Naval Training Equipment Center.
- Obermayer, R.W., & Vreuls, D. (1987, December). Combat-ready crew performance measurement system: Phase II measurement system requirements (AFHRL-TR-74-108 (III), AD-B005 519L). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory.
- Rasinski, J.E., Rabeni, J.J., Holck, E.K., & Pierce, B. (1983). A-10 operational flight trainer (OFT) Phase I follow-on operational test and evaluation (AFOTEC Project 81-AFTEC-213). Kirtland AFB, NM: HQ Air Force Operational Test and Evaluation Center.
- Schmid, R.F., Gibbons, A.S., Jacobs, R., Faust, G.W., & Moore, C.T. (1978). Recommendations for the F-16 performance measurement system (F-16 Development Report No. 14). San Diego, CA: Courseware Inc.
- TAC Manual 51-50. (1982, March - 1984, January). Flying training (Vol. II, A-10 Pilot Training, Vol. VII, F-15 Aircrew Training, Vol. VIII, F-16 Aircrew Training). Langley AFB, VA: HQ TAC.
- TAC Regulation 50-31. (1983, January). Training records and performance evaluations in formal flying training programs. Langley AFB, VA: HQ TAC.

TAC Regulation 60-2. (1983, February). Aircrew standardization/evaluation program (Vol. I-IX). Langley AFB, VA: HQ TAC.

Waag, W.L. (1983). Performance measurement requirements in the design of instructional support systems for operational training simulators. In Proceedings of the Royal Aeronautical Society Symposium on Training Simulators - A User's Forum. London, England.

Waag, W.L. (1987, in publication). Instructional support features for major training devices (Technical Report of UTP-2 of The Technical Cooperation Program) Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Waag, W.L., & Hubbard, D.C. (1984). The measurement of C-5A aircrew performance. In Proceedings of Psychology in the DOD Symposium. Colorado Springs, CO: US Air Force Academy.

Waag, W.L., & Knoop, P. (1977). Planning for aircrew performance measurement R&D: US Air Force. In Proceedings of Naval Personnel Research and Development Center (NPRDC) Symposium on Productivity Enhancement: Personnel Performance Assessment in Navy Systems. San Diego, CA: NPRDC.

APPENDIX A: INTERVIEW PROTOCOL

User Performance Measurement System Interview Guide

I. User Job Responsibility:

A. Job Function.

B. Responsible for What Personnel/Agencies:

1. Specific areas of responsibilities for each of these personnel/agencies.
2. Function of these areas (i.e., relate areas to conduct or management of training).

C. Responsible to What Personnel/Agencies:

1. Type of training information provided to those personnel agencies.
2. Format of that training information.
3. Use of that training information.
4. Frequency with which that training information is required.

II. Current Sources: For each task or category of tasks, identify:

- A. Source and format of performance information.
- B. Amount of information (feedback).
- C. Validity of information.
- D. Reliability of information.
- E. Sensitivity of information.
- F. Suitability of information.

III. Performance Measurement Enhancements:

A. Based on information from parts II and III, identify deficiencies in current measurement information flow.

B. Optional: Discuss user suggestions for remedial action.

APPENDIX B: INTERVIEW SCHEDULES

Unit	Offices Interviewed
405 TTM	Wing Training Wing Simulator Training Wing/Squadron Standardization-Evaluation Wing/Squadron Weapons Wing/Squadron Scheduling Academics/Training Development Squadron Commanders/OPS Officers Squadron Flight Commanders Instructor Pilots/Student Aircrews
58 TTM	Wing Training Wing Simulator Training Wing Weapons Instructor Pilots
355 TTM	Wing Training Wing Simulator Training Wing Weapons Wing Standardization-Evaluation Academics/Training Development Instructor Pilots
474 TFW	Wing/Squadron Training Wing/Squadron Weapons Wing/Squadron Standardization-Evaluation Wing/Squadron Scheduling Wing/Squadron Checkered Flag Wing Intelligence Wing Simulator Training Wing Flight Records Squadron OPS/Flight Commanders
49 TFW	Wing/Squadron Training Wing/Squadron Weapons Wing/Squadron Scheduling Wing Simulator Training Wing Flight Records Squadron Flight Commanders
HQ TAC	Weapons (DOOW) Simulator Training (DOTS) Formal Training (DOTF-10,15,16) Training Development (4444) Standardization-Evaluation (DOVF) Aggressors/Red Flag (DOOO) Flight Records (DOXBA)